Classroom-based Assessment in L2 Contexts

Classroom-based Assessment in L2 Contexts

Edited by

Dina Tsagari

Cambridge
Scholars
Publishing

Classroom-based Assessment in L2 Contexts

Edited by Dina Tsagari

This book first published 2016

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2016 by Dina Tsagari and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-9102-9
ISBN (13): 978-1-4438-9102-8

TABLE OF CONTENTS

INTRODUCTION 1
Dina Tsagari

PART I: LANGUAGE ASSESSMENT LITERACY

CHAPTER ONE 8
Assessment Literacy for Language Teachers
Antony Green

CHAPTER TWO 30
Language Assessment Literacy in English Teacher Education in Hungary
Ildikó Csépes

CHAPTER THREE 54
A Case of Training University Teachers in Developing and Validating Classroom Reading Test Tasks
Olga Kvasova

CHAPTER FOUR 75
Assessment as Technology of Power: Shaping Literacy and Literate Identities in a Public Elementary School
Stavroula Kontovourki

PART II: WASHBACK IN THE L2 CLASSROOM

CHAPTER FIVE 96
Examining the Quality and Quantity of the Washback Effect of the FCE Test in Greece
Lambrini Loumbourdi

CHAPTER SIX 118
A Washback Study of the Teaching Practices used in EFL Multi-Exam Preparation Classes in Greece
Irini Papakammenou

CHAPTER SEVEN 138
The Washback Effect of Peer Assessment on EFL Learners in Cyprus
Eleni Meletiadou and Dina Tsagari

CHAPTER EIGHT 161
Washback of an Oral Exam on Teaching and Learning in German Middle Schools
Veronika Froehlich

PART III: FORMATIVE ASSESSMENT

CHAPTER NINE 184
Assessment for Autonomy, Learning, and Learner Motivation: Fostering Learner Identities
Terry Lamb and Sabine Little

CHAPTER TEN 207
Towards an Alternative Approach to Oral Performance Assessment in English as a Lingua Franca
Androniki Kouvdou

CHAPTER ELEVEN 229
Process Writing and Peer-Assessment in Teacher Education
Claudia Mewald

CHAPTER TWELVE... 254
EFL Primary School Teachers' Understanding of FA in Cyprus
George Michaeloudes and Dina Tsagari

CHAPTER THIRTEEN 273
Formative Assessment Practices in Private EFL classes in Cyprus
Dina Tsagari and Andrie Kontozi

PART IV: ASSESSING L2 SKILLS AND TEACHING MATERIALS

CHAPTER FOURTEEN 298
What is Fluency? Reading a Common Understanding
Hatice Çelebi and Gaele Macfarlane

CHAPTER FIFTEEN 330
Assessing Spanish in a Communicative School Context: The Case of Teacher-Made Tests in Austria
Barbara Hinger

CHAPTER SIXTEEN 356
Integrating Corpus Linguistics and Classroom-based Assessment: Evidence from Young Learners’ Written Corpora
Trisevgeni Liontou and Dina Tsagari

CHAPTER SEVENTEEN 383
Linking Materials to the CEFR
Karin Vogt

CONTRIBUTORS 405

INTRODUCTION

DINA TSAGARI

It is refreshing that the field of language testing and assessment has recognised the importance and underlying theoretical and practical underpinnings of classroom-based language assessment (CBLA), an area that is gradually coming into its own. For many years, CBLA has been mainly viewed as a tool to record student achievement through the types of items and tasks employed in traditional large-scale testing. In reality, however, CBLA is used to support learning and inform teaching. With the growing awareness of assessment activities internal to the classroom and managed by teachers, in addition to research designed in mainstream classrooms (Black and Wiliam 1998, 2009), CBLA has become an emerging paradigm of its own, with a focus on language learning and an evolving research agenda (Purpura and Turner, forthcoming; Rea-Dickins 2008; Turner 2012; Tsagari and Csépes 2011). Research in CBLA has critically expanded, and new, exciting, and uncharted avenues where theory and research meet in a dialectic and informative relationship have appeared.

Against this background, this book addresses issues that promote the notion of CBLA for the academic community and beyond in language research, teaching, and any related areas. The book explores recent thinking and research on CBLA in the field of language testing, assessment, and general education based on the collection of theoretical and research papers presented over the recent Classroom-based Language Assessment SIG Symposia organised in Cyprus and pre-conference EALTA workshops in various countries around Europe.[1]

The book comprises 17 chapters organised in four parts that follow a logical structure integrated around the following central themes:

Part I: Language Assessment Literacy (4 papers)
Part II: Washback in the L2 classroom (4 papers)
Part III: Formative Assessment (5 papers)

---

[1] See http://www.ealta.eu.org/resources.htm.

Part IV: Assessing L2 skills and teaching materials (4 papers)

The chapters examine the issues mentioned above through a multidisciplinary and multifaceted approach, and intertwine them from different perspectives. Contributors to the volume also address the issues from a theoretical as well as an empirical point of view. Studies of both cross-sectional and longitudinal nature are included, as well as studies conducted with young and adult test takers in either high- or low-stake environments.

The first chapter of Part I by Antony Green lays the ground for the three research-oriented chapters following. In his chapter, Green defines the notion of language "assessment literacy" (LAL) as part of a teacher's CBLA repertoire and professional development, and addresses the teachers' levels of assessment literacy in order to make effective use of assessment information in meeting their CBLA needs. The next three chapters report on research conducted in LAL in various countries. In the first of these chapters, Ildikó Csépes presents the results from a survey on English teacher trainees' perceptions of their LAL after their graduation from an MA programme in the Hungarian context. he study also discusses contextual factors that play a role in shaping the curriculum of teacher training programmes in language testing and assessment.

In the next chapter, Olga Kvasova investigates Ukrainian foreign university teachers' ability to construct and validate reading test items/tasks for classroom use. The author describes in detail a training module that was developed and carried out in several in-service training contexts in the country. Other than the effectiveness of the designed module, the results also stress the important role of collaboration within teams of test designers, and the role of testees to improve assessment tools that can provide accurate feedback. In the final chapter of this part of the book written by Dr Stavroula Kontovourki, literacy assessment is approached as a technology of power whereby particular meanings of literacy and literate identities are being regulated and produced. The chapter draws from an ethnographic study of children's performances of literacy in a third-grade public school classroom in a US city. Tracing how the language of assessment cuts across CBLA events, the hybridity of schooled literacy is discussed and questions are raised regarding the less visible ways in which assessment shapes the literacy performances of readers and writers in public school classrooms.

The next four chapters comprising Part II report research findings within the field of "washback" – the influence that testing has on teaching and learning (Alderson and Wall 1993) – an important aspect of CBLA.

The first chapter, written by Lambrini Loumbourdi, focuses on the role of washback and the relationship between the quality of a test, the impact of educational and social beliefs, and the issue of exam preparation within the context of FCE (First Certificate in English) administered by the Cambridge English Language Assessment in Greece (see also Tsagari 2009). The chapter offers a range of pedagogical implications addressing the professional development of teachers in the context of exam washback. In the same educational context, Irini Papakammenou reports on her washback study conducted in private multi-exam preparation classes where a variety of exams are taught in the same class, presenting an interesting pedagogical ecology. The chapter discusses the nature and type of methodology teachers use in these classes and the extent to which their approaches are influenced by the nature and requirements of the exams. The chapter emphasises the need to draw a distinction among "methods", "activities", and "tasks" when studying teachers' methodology in washback studies, and offers various stakeholders insights into the workings of washback relevant to their needs.

In the third chapter, Eleni Meletiadou and Dina Tsagari investigate whether peer assessment (PA) as a form of CBLA, increasingly being adopted in secondary education settings, can induce positive washback on the participants' attitudes towards EFL (English as a Foreign Language) writing in adolescent classes in Cyprus. The results showed positive washback on learners' attitudes towards PA of writing. The chapter offers pedagogical recommendations with regard to the use of PA mechanisms that foster positive washback on learners who can be assisted in their effort to become more goal-oriented and consequently more self-directed. In the final chapter of this part, Veronika Froehlich reports on the impact of an oral exam (the EuroCom), introduced in German middle schools following the educational reforms in 2004, on teaching and learning. Despite expectations that EuroCom would have a positive washback effect, the results showed that the effects on teaching and learning are complex and induced by a number of factors such as teacher education and motivation, teacher and learner understanding of exam requirements, the teachers' ability and willingness to adapt their teaching content and styles to support positive washback, the publishers' production of appropriate material, the availability of teacher training seminars, and student motivation. The chapter offers important implications for those involved in student preparation for high-stakes exams.

The next four chapters, Part III, focus on formative assessment (assessment *for* learning rather than summative assessment, or assessment *of* learning). Formative assessment (FA) refers to a range of formal and informal assessment procedures employed by teachers as part of their classroom assessment procedures in order to identify students' weaknesses and strengths and define future learning goals to improve student learning (Rea-Dickins 2004). In the first chapter of this part, Terry Lamp and Sabine Little explore FA as a way of providing justification for assessing autonomy. The chapter argues that, just as language learning can be enhanced by assessment for learning, so can learner autonomy be enhanced by assessment for *autonomy*. The chapter presents research and classroom practice in foreign language learning in secondary schools in England, and considers issues for teacher development and implications for classroom practice.

In the next chapter, Androniki Kouvdou discusses whether alternative assessment is appropriate for the assessment of oral performance skills in Greek state school classes, a teaching context where the principles of English as a Lingua Franca (ELF) are highly valued. The chapter narrows the existing gap between CBLA and ELF by providing substantial evidence for the potential effectiveness of systematic and focused observation as an assessment tool. The next chapter by Claudia Mewald looks into ongoing research at the College of Teacher Education in Lower Austria. There, process writing and peer assessment based on a negotiated criterion-referenced assessment scale are implemented in literature courses where English is used as a foreign language. The chapter concludes with pedagogical implications for language educators in the current context and beyond.

George Michaeloudes and Dina Tsagari report on state EFL teachers' background knowledge of formative assessment (FA) in the primary school context in Cyprus in their chapter. The chapter examines teachers' understanding of FA using qualitative and quantitative approaches for the analysis of the questionnaire data collected. The chapter discusses the findings in respect of the existing literature, and concludes with suggestions for further research in FA for CBLA purposes. In the final chapter of this part, Dina Tsagari and Andrie Kontozi also report on the nature of FA practices occurring in the private classes of EFL teachers in Cyprus. In particular, the study examines the nature of interactions taking place during informal FA procedures, and demonstrates how these are operationalised during teaching. To do so, the study was based on an in-depth analysis of recorded classroom observations. The chapter concludes with suggestions for teacher training, aiming to help practitioners become assessment literate in FA and, consequently, foster successful language learning.

The final four chapters focus on issues relating to the assessment of various L2 skills and teaching materials. For example, Hatice Çelebi and Gaele Macfarlane examine how the reliability of a classroom-based speaking exam can increase in summative classroom-based and institutional proficiency speaking exams. Their chapter presents the findings and the implications of the process the authors went through to create classroom-based rubrics for the assessment of oral proficiency. In the next chapter, Barbara Hinger analyses written classroom-based Spanish exams in the Austrian upper secondary school context following a descriptive-exploratory research approach. The analysis includes qualitative and quantitative approaches, and the chapter offers insights into the application of communicative language test principles in the school context.

In the third chapter of this part, Trisevgeni Liontou and Dina Tsagari report on a longitudinal study that aimed at assessing EFL young learners' writing proficiency as a function of grade level in Greek primary classes. The results showed statistically significant differences between the linguistic features identified in the essays produced by young learners at different levels of language competence. The chapter concludes by providing practical guidance to EFL teachers as to the kind of linguistic strategies young learners develop as a function of their grade level, and offers suggestions for designing classroom curricula and classroom-based writing assessment activities. In the final chapter, Karin Vogt looks closely into the relationship between foreign language teaching materials and the CEFR (Council of Europe 2001). The CEFR has been embraced by educational authorities far and wide exploring whether the teaching material used and competences to be acquired are linked to the CEFR level to be achieved in the foreign language classrooms in Germany. The chapter reports on a mixed methods research project which aimed at creating teaching materials for EFL based on adapted CEFR descriptors, and explores further ways to strengthen the link between teaching materials and the CEFR.

Overall, this edited volume with its international scope aspires to become a ground-breaking resource, bringing together in a balanced relationship the fields of education and second-language testing and assessment. As such, the book brings together scholars who are at the cutting edge of the ever-expanding fields mentioned above, critically addresses the issues involved in CBLA, and broadens the approaches undertaken to date. Given its scope and nature, the volume constitutes an important reference source and reading material in the field of language testing and assessment at universities, and a point of reference in pre- and in-service teacher training in general and language education.

Works Cited

Alderson, J. Charles and Dianne Wall. 1993. “Does Washback Exist?” *Applied Linguistics* 14:115–29.

Black, Paul, and Dylan Wiliam. 2009. “Developing the Theory of Formative Assessment.” *Educational Assessment, Evaluation, and Accountability* 21:5–31.

Black, Paul, and Dylan William. 1998. “Assessment and Classroom Learning.” *Assessment in Education* 5:7–74.

Council of Europe. 2001. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: Cambridge University Press.

Purpura, James E. and Carolyn E. Turner. forthcoming. *Learning-oriented Assessment in Language Classrooms: Using Assessment to Gauge and Promote Language Learning*. New York, NY: Routledge, Taylor & Francis.

Rea-Dickins, Pauline. 2004. “Understanding Teachers as Agents of Assessment.” *Language Testing* 21(3):249–58.

Rea-Dickins, Pauline. 2008. “Classroom-based Assessment.” In *Encyclopedia of Language and Education*, edited by Elana Shohamy and Nancy H. Hornberger, 1–15. USA: Springer.

Tsagari, Dina and Ildikó Csépes (eds.) 2011. *Classroom-based Language Assessment.* Frankfurt am Main: Peter Lang GmbH.

Tsagari, Dina. 2009. *The Complexity of Test Washback*. Frankfurt: Peter Lang.

Turner, E. Carolyn. 2012. “Classroom Assessment.” In *The Routledge Handbook of Language Testing*, edited by Glenn Fulcher and Fred Davidson, 65–78. London and New York: Routledge.

PART I:

LANGUAGE ASSESSMENT LITERACY

CHAPTER ONE

ASSESSMENT LITERACY FOR LANGUAGE TEACHER

ANTHONY GREEN

Language teachers need to be well informed about the principles and practice of assessment. Unfortunately, current levels of training are inadequate to meet their requirements. Firstly, assessment is given too little attention in teacher education. Secondly, the expansion of language assessment internationally and changes in the roles that teachers are expected to play mean that established professional development models need to be transformed. Although the need for reform is widely accepted, there is little consensus over how training should be changed. Some take the view that innovative classroom assessment practices can be accommodated within traditional approaches to language testing; others argue that new theoretical approaches are needed. This chapter introduces the concept of learning-oriented language assessment (LOLA) as a basis for reconciling different purposes for assessment and enhancing the coherence of educational systems. LOLA requires that all assessment tasks should reflect curriculum objectives, involve opportunities for language learning, offer feedback, and engage learners in the assessment process.

1. The origins of assessment literacy

The term *assessment literacy* can be traced back to a short article written by Rick Stiggins (1991) in which he lamented a widespread inability to interpret assessment information and the failure of specialists in educational measurement to improve the situation through effective communication and training.

Stiggins gave examples of the importance of understanding the uses and implications of assessment data on the part of students, parents, members of school boards, legislators, and public officials. He expressed his greatest concern over the shortcomings of teacher training, noting that "the majority of teacher education programmes require no training in assessment for graduation" (Stiggins 1991, 535). He suggested that training teachers in assessment was essential to effective student learning, and called for a redirection of research efforts towards assessment in the classroom rather than standardised testing. The need to shift the focus of attention towards teacher education and the role of assessment in the classroom has been a persistent theme for advocates of assessment literacy.

While Stiggins (1991) was writing about general education in US schools, recent years have seen similar concerns being voiced about levels of assessment literacy in international language education. As in Stiggins' article, the difficulties that a wide range of groups experience in interpreting and using assessment data have been discussed. These groups include university admission officers, policy makers, and the general public. However, the topic attracting the greatest attention has been "the assessment skills and understandings currently perceived as vital for conducting language assessment in educational settings" (Inbar-Lourie 2008, 387): the core competences in assessment required for effective language teaching.

2. Language testing and teacher training

In some ways, recent debates about assessment literacy can be viewed as simply the latest expression of one of the founding aims of language testing as a branch of applied linguistics: to enlighten language teachers about the design and use of tests. Robert Lado, in a book widely cited as the starting point for language testing as a distinct discipline, addressed himself "primarily" to "teachers of foreign languages and to teachers of English as a foreign language" (Lado 1961, vii).

Since Lado, there has been a proliferation of publications about testing, meeting the growing demand from teacher education programmes (see Davies 2008). Nonetheless, a recurrent complaint from language testing specialists has been that the topic tends to be neglected or marginalised in teacher training. There is a perception that even programmes aimed at more advanced students and experienced teachers fail to provide sufficient input on assessment. Taylor (2009, 23), for example, complained that graduate programmes for language teachers "typically devote little time or attention to assessment theory and practice, perhaps just a short (often optional) module".

3. What do courses in language assessment cover?

In response to concerns about a deficit in teacher assessment literacy, a number of attempts have been made to better define and explicate the range of competences in assessment that teachers need to develop. Shortly before Stiggins published *Assessment Literacy*, the American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and the National Education Association (NEA) (1990) produced a set of standards for teacher competence in the educational assessment of students (see Appendix A).

These standards have exerted a strong influence on discussions of assessment literacy in language education and have been widely quoted. However, as Stiggins (1991) cautioned, the meaning of assessment literacy varies according to context. It is unlikely that standards derived from the US national education system can be applied to international language education without being reshaped in some way. Bailey and Brown (1996, 250), therefore, suggested that "it behoves us, as a profession, to consider drafting or adapting a similar set of standards dealing specifically with language assessment".

In setting out the core assessment competences required by language teachers, there have been efforts to both ascertain what is generally covered in teacher training courses and evaluate how well teachers are able to apply what they have learned.

4. Surveys of language testing course instructors

One strand of research has involved surveys of language assessment course content. Bailey and Brown carried out two global surveys (with 84 and 97 respondents, respectively) concerning topics covered by graduate courses (Bailey and Brown 1996, Brown and Bailey 2008). Although there was some variation in content, relatively little appeared to have changed over the intervening decade. Courses consistently featured such core topic areas as validity theory, reliability, measurement error, statistics for test and item analysis, the critique and analysis of test content, and item-writing skills.

Jin (2010), who carried out a parallel survey of 86 instructors in China, found similar evidence of the enduring appeal for tutors of established course constituents. Test validity and reliability, principles and practice of item writing, and the construction of (multiple-choice) test questions again featured among the most commonly taught topics.

Jeong (2013) also followed Brown and Bailey (2008) in her survey of 140 language testing course leaders around the world. Her goal was to involve instructors who lacked specialist language assessment qualifications in order to correct a bias towards language testing specialists in Brown and Bailey's data (their respondents had been recruited via a language testing discussion list). Although her data suggested that courses taught by both groups tended to follow similar patterns, it also pointed to a greater diversity of provision. Non-specialists tended to give greater attention to classroom assessment topics while specialists spent more time on statistics.

5. Tests of assessment literacy

In addition to the surveys of language testing courses, another, less developed, strand of research has proposed the use of tests of assessment literacy as a means of collecting more direct evidence of the extent to which language teachers are aware of key testing concepts.

Plake and Impara (1993) used the 1990 AFT et al. standards as the basis for a 35-item test of teacher assessment literacy in general education in the United States. The test items were developed as application questions that would be relevant to teachers' experiences. Their national survey of 555 teachers revealed that teachers spent up to 50 percent of their time on assessment-related activities. However, the average score on the test was just 23 out of 35, pointing to deficits in most of the areas tested.

Some researchers have taken up Plake and Impara's (1993) approach and applied it to language testing. Newfields (2006) constructed a test of language teacher assessment literacy designed for self-diagnosis. His test covers four broad topic areas: Terminology, Procedures, Test Interpretation, and Assessment Ethics. However, the majority of the questions focus on the use and interpretation of statistics. Kaftandijeva (2008) further refined Newfields' test to create a 27-item True/False test for online administration. Unfortunately, the reduction in length comes at the cost of an even more restricted coverage of topics. Although testing the assessment literacy of teachers appears a promising theme, no results are available showing how well teachers perform on either the Newfields (2006) or Kaftandijeva (2008) measures.

6. Tests made by teachers

Other researchers have taken a more qualitative observational approach to teachers' ability to apply language testing concepts. Coniam (2009) noted findings from previous studies showed that tests made by teachers tended to be of poor quality, “were too difficult or too easy; … measured content that had not been taught in class or not specified in the syllabus… [and] did not show what students had actually achieved” (227). His study explored the effects of a basic training course in assessment (provided as part of a graduate course in English Language Teaching) on the quality of teacher-produced tests.

Although teachers were able to improve their material by editing, reviewing, and redrafting, even after passing through the recommended stages of test construction, the resulting tests generally failed to satisfy basic quality criteria. In addition to insufficient training, Coniam’s (2009) informants reported that practical constraints had a negative impact on their work: lack of time, resources, and institutional support restricted the attention they could give to developing effective assessment and improving their own knowledge and skills.

7. Changing needs

Adding new urgency to the long-standing concern about the lack of attention to language testing in teacher training, observers have pointed to a recent expansion in the role that teachers are expected to play in assessing language learners – including testing for high-stakes purposes (those with significant implications for the life chances of test takers such as access to higher education and employment opportunities). At the same time, new questions have been raised about the adequacy of what little is on offer.

In addition to the need to understand and develop tests, which is covered by Lado (1961) and his successors, commentators have noted a shift in the focus in educational policy-making away from inputs to learning (curriculum) towards outcomes from learning (as evidenced by assessment results) (Cumming 2009). The burden for generating evidence of outcomes has mainly fallen on teachers, greatly increasing both the amount of assessment they are asked to carry out and the importance placed on it.

8. Teacher assessment for accountability

Fulcher (2012) associated this expansion in teacher responsibility for assessment with the adoption of accountability systems and the growing influence of frameworks such as the Common European Framework for Languages (CEFR) (Council of Europe 2001). The CEFR defines learning outcomes at different levels of proficiency in concrete, practical terms such as "Can write personal letters describing experiences, feelings and events in some detail" (Council of Europe 2001, 83). Teachers are expected to develop procedures for assessing whether learners meet the intended standards: procedures which may not be limited to the kinds of test described by Lado (1961) but may include alternatives such as projects and learner portfolios.

9. Teacher assessment for learning

Alongside the greater role of teachers in judging and reporting learning outcomes (summative assessment), assessment for learning (Black and Wiliam 1998) and allied movements offer policy makers an attractive vision in which teachers can use assessment to gain insights into learning processes and so guide and enhance learning processes in the classroom (formative assessment). They may also be expected to support learners in assessing themselves and other learners.

Although there is clearly great variation related to national and institutional cultures, teachers are now often required to manage at least three distinct assessment functions, each of which involves a different set of assumptions about assessment (Cumming 2009). These include:

i) preparing students to face high-stakes tests designed to discriminate between more and less successful learners.
ii) designing and administering summative assessments capable of indicating whether or not their students have met explicit curriculum standards.
iii) implementing formative assessment policies intended to cultivate more successful language learning.

10. Limitations of traditional language testing courses

If the role of teachers is changing, this suggests a need not only to increase the amount of training on offer but also to change the nature of courses in language assessment to meet new requirements. One symptom of a change in direction is the growing use of the term "language assessment" rather than the more established "language testing". Although the distinction is imprecise and inconsistent, the use of "assessment" generally implies a wider range of techniques for investigating learners' language abilities. It embraces such alternatives as learner portfolios and self-assessment in addition to tests. It is sometimes also used to suggest a preference for, or at least an acknowledgement of, theories that challenge the established premises of language testing. Disagreements about the urgency and nature of reform mean that choices about the content of courses are now more controversial and contested than they were in previous decades.

Descriptions of current practices of the kind offered by Bailey and Brown (1996, Brown and Bailey 2008) can serve to inform providers and disseminate current practices. On the other hand, they offer only relatively established, conservative patterns for the design of new courses. They may allow us to take stock of what is currently on offer, but ascertaining what *is* taught may tell us relatively little about what *should be* taught to meet teachers' needs in rapidly changing educational environments. The new millennium has seen an increased questioning of the standard language testing course template that they describe.

11. The focus on testing in coursebooks

Several commentators have called attention to the gap between the content of language testing courses and teachers' assessment responsibilities. Historically, language testing books have tended to concentrate on approaches to test development and quality assurance developed for use in large-scale testing programmes. Inbar-Lourie (2008) pointed out that many books give little or no attention to classroom assessment or even to summative achievement testing in schools. In other words, they partly address the first two of the assessment functions outlined above. They typically inform teachers about the design and evaluation of test instruments but tend not to offer much guidance on test preparation or on the development of alternative assessments based on curriculum standards. They offer little or no direction on the third assessment function: assessment to support teaching and learning processes.

Brindley (2001) was among the first to question the relevance of traditional language testing course content to teachers' needs. He suggested that because much of the assessment carried out by teachers did not involve traditional tests, there was relatively little scope for teachers to apply the knowledge gained from language testing courses.

Davies (2008) identified the first coursebook in the field to be concerned primarily with classroom assessment to be by Genesee and Upshur (1996). However, this book featured as required reading on just three of the courses surveyed by Brown and Bailey (2008). The two most popular textbooks were by Bachman and Palmer (2006) and Brown (1996), each used on 18 courses. Both books focus mainly on testing viewed as a form of measurement supported by statistical analyses. The five most popular text books on the survey were all first published in the 1980s or 1990s.

The tendency to rely on established sources was even more marked in the Chinese context covered by Jin's (2010) study. Bachman (1990) and Heaton (1988) were the two most popular books (mentioned as core text books by 24 and 19 respondents, respectively). All the titles of the 12 most popular books listed in the article featured the words "test" or "testing" rather than "assessment", and only two were published after 2000. Genesee and Upshur (1996) did not feature at all.

12. The technical focus of language testing courses

A second line of criticism directed at language testing coursebooks concerns the overly technical nature of their content. One aspect of this is accessibility: the focus on large-scale testing brings with it specialised jargon, graphs, and statistical formulas that risk alienating many language teachers and teacher trainers.

There is also a concern that training may concentrate too much on *how* – and too little on *why* – assessments are made. Courses of the kind described as typical by Brown and Bailey (2008) tend to fall into what Davies (2008, 328) in his historical overview of LT coursebooks termed a "knowledge + skills" training model: one that provides a background in measurement with practical procedures for test analysis and construction. The missing element in this training model, Davies suggested, is "principles", or scrutiny of the beliefs and assumptions that inform testing practices and the consequences that follow from their use.

Davies reviewed four of the five most popular books featured in the Brown and Bailey (2008) survey. Although he judged that the more recent texts were filling the gap that he identified, he considered that just two of the four dealt with principles to any significant extent.

13. Effects of the neglect of classroom assessment

The neglect of classroom assessment in training may inhibit the use of innovative techniques or contribute to the generally poor quality of classroom assessment procedures and tools developed by teachers. Evidence, mostly from general education, suggests that teachers often struggle to use assessment effectively to support learning. They rely on familiar testing techniques and emphasise the summative role of assessment in grading learners rather than its formative potential as a means of improving performance (Brindley 2001).

Two linked pan-European surveys of language teachers involving a total of over 1,500 respondents (Hasselgreen, Carlsen, and Helness 2004; Vogt and Tsagari 2014) found that the majority of teachers had received either little or no training in classroom assessment. They also lacked the ability to critically evaluate the assessments they used. Often relying on their own experience as test takers or learning on the job from colleagues, teachers were generally much more confident about using traditional testing than alternatives such as portfolio assessment, self-assessment, or peer assessment.

Case studies of teachers' practices also suggest that teachers find classroom assessment challenging. They face problems in reaching consistent judgements about performance, accepting and interpreting curriculum standards and scoring criteria, and understanding what classroom data might reveal about learners' language development (Davison 2007; Leung and Mohan 2004; Rea-Dickins and Gardner 2000).

14. Dimensions of assessment literacy

Given that assessment is called on to fulfil a wide range of purposes, it has been suggested that different social groups may require different kinds of assessment literacy or different blends of knowledge, skills, and principles. Stiggins (1991) originally proposed three levels of assessment literacy: a *functional* level required by public officials and others who need to use assessment results; a *practical* level needed by teachers and other educators who not only use assessment data but also need to generate it; and an *advanced* level required by specialists in educational measurement who generate data for others to use. He also suggested that assessment literacy was multi-dimensional, and the balance of skills and knowledge needed by each of these groups would vary according to context.

In line with the criticisms of traditional courses noted above, there have been calls to redesign courses for language teachers: to shift attention towards classroom assessment rather than tests. Brindley (2001, 131) suggested that "since most teachers are not engaged in the construction of formal tests, there are strong arguments for placing the emphasis in professional development – at least initially – on the role of assessment in the learning process rather than on theoretical and statistical issues in testing." Training courses for teachers should "begin with a focus on curriculum-related assessment" (Brindley 2001, 129). They should also include such alternatives to testing as the use of "observation schedules, portfolios, conferences, project work, journals, self-assessment techniques and progress and achievement profiles" (Brindley 2001, 130).

15. Levels of assessment literacy

Taking up the suggestion of levels of assessment literacy from Stiggins (1991) and applying it to language assessment, Taylor (2013) proposed three assessment literacy levels (*core*, *intermediary*, and *peripheral*) that could be distinguished according to eight areas (see Appendix A). She recommended that teachers would need to have only a relatively basic awareness in three topic areas that are core to the traditional language-testing courses surveyed by Brown and Bailey (2008): *knowledge of theory*, *principles, and concepts, scores, and decision making*. In place of this established core, she suggested that during teacher training, priority should be given to issues of *language pedagogy*, *sociocultural values*, *personal beliefs/attitudes*, and *technical skills*.

Taylor (2013) acknowledged that assessment literacy was multi-dimensional and that teachers (who she placed at the intermediary level) might not require all of the skills needed by testing specialists. However, she also apparently assumed that researchers and professionals in language testing represented and defined the highest level of assessment wisdom across all of the dimensions in her framework. This would seem to imply that although some adjustment or rebalancing of course content may be required to take account of teacher needs, specialists in language testing (researchers and test makers), forming the inner circle of assessment literates, are best placed to determine the content of training courses.

Jeong (2013) illustrates how this somewhat condescending assumption may shape the interpretation of data in assessment literacy research. She took the discrepancies she found between courses taught by language-testing specialists and non-specialists as evidence of a deficit in expertise on the part of the non-specialists. She concluded that better communication of insider expertise was necessary: "it is the role of the language testing community to make the field approachable to others" (357). However, an alternative interpretation might be that the language- testing specialists were less sensitive than non-specialists to their trainee teachers' need for guidance on classroom assessment.

There is some evidence from other studies that, contrary to her interpretation, Jeong's (2013) non-specialist informants better reflected teachers' concerns. Malone (2013, 343), reporting on the evaluation of assessment training materials by LT specialists and language teachers, concluded that "what interests language testing experts may not be of interest to language instructors”. This suggests the possibility of a more dialogic approach to training that acknowledges and takes account of the educational cultures and experiences the trainees bring with them.

Fulcher’s (2012) global survey also targeted teachers. He asked 278 teachers what they considered should be covered by language assessment courses. Respondents were critical of recognised coursebooks for ignoring or downplaying the social purposes of assessments, historical precedents, fairness, and the ethics of testing practices. They also called for a more balanced treatment of classroom and large-scale assessment.

Although both changes over time and the needs of different stakeholders have been noticed, it seems surprising that in discussions of language assessment courses for teachers, comparatively little attention has been given to cultural differences. There is certainly considerable variation in the relative weight given to assessment by teachers and to large-scale testing in different parts of the world. In some educational systems, teachers have considerable autonomy in deciding how and when they will conduct assessments, and their decisions have important consequences for learners. In others, they have little scope even to make use of formative assessment data to adjust the rate or content of teaching.

16. Continuity between language testing and classroom assessment

While calling for greater attention to classroom assessment, many language testing specialists continue to subscribe to a model for training that centres around the established language testing traditions. Prescriptions for innovations in course design typically include incorporating classroom achievement testing techniques and criterion referencing (judging performance in relation to a predetermined standard rather than to the performance of other learners) within the established knowledge and skills format. They enrich the format by exploring the ethics and social consequences of test use.

From this perspective, new methods associated with classroom assessment can be assimilated and evaluated in relation to the same criteria as large-scale tests. An example can be found in Brown and Hudson’s (1998) paper which argued that classroom assessments should be judged according to the traditional standards for test validity. As one of Jeong’s informants expressed it, “Validity is validity, and reliability is reliability” (Jeong 2013, 353) regardless of whether the context is classroom assessment or large-scale testing.

17. Contrasting paradigms and irreconcilable differences

This relatively conservative model of assessment literacy has been challenged in recent years. It has been argued that the field is more diverse and the nature of assessment expertise more contested than the traditional model suggests. New conceptualisations of language assessment have begun to emerge that question the epistemological assumptions of the field. It has been suggested that distinct approaches to assessment qualities are required to take account of classroom assessment purposes (Moss and Brookhart 2012, Shepard 2000): validity and reliability may need to be reconceptualised or different qualities may need to be considered.

The emergence of alternative theoretical approaches to complement classroom assessment techniques make it possible to offer assessment courses for language teachers that owe little to the language testing tradition. Taking what I will call, after Scarino (2013), a *contrasting paradigms* approach, Inbar-Lourie (2008, 389) argued for a wholesale "reformulation of the competencies needed for conducting assessment in the educational context." Following Shepard (2000) and Lynch (2001), Inbar-Lourie (2008) suggested that courses for teachers should embrace a sociocultural paradigm, or "assessment culture", for classroom assessment in preference to the familiar measurement paradigm, or "testing culture", reflected in traditional language testing courses.

18. Assessment cultures

In contrast to the test-centred definition of assessment literacy proposed by Fulcher (2012) in Appendix A, Inbar-Lourie (2008) preferred the following: "the ability to understand, analyze, and apply information on student performance to improve instruction" (Falsgraf, 2006, 6 cited in Inbar-Lourie 2008, 390). In this definition, it is teaching rather than testing that is the central concern. Approaches that draw on Vygotskian sociocultural theory, such as assessment for learning and dynamic assessment, are contrasted with and preferred over psychometrics as a basis for evaluating assessment design and use.

In an assessment culture, the role of the teacher involves "formulating and scaffolding learning on the basis of on-going feedback from internal and external assessment sources" (Inbar-Lourie 2008, 388). Precedence is given to forms of assessment carried out by teachers and the learners themselves in the course of regular classes. Key outcomes are not scores or grades that are readily amenable to statistical analysis but insights into the learning process that can guide development. Inbar-Lourie (2008) introduced the term *language* assessment literacy to emphasise the value of training teachers in how languages are learned and how learners are motivated to learn because these can support effective scaffolding and feedback processes.

Proponents of sociocultural approaches have suggested that embracing an assessment culture implies training in the use of different quality standards to those found in the established LT coursebooks. Assessment methods should not be judged according to how efficiently they discriminate between those who “know the answers” and those who do not (as in traditional testing). They should be evaluated according to how much they reveal about why or how well a concept is known or understood – or how close a student comes to successful independent performance. Assessment data gathered through interaction in the classroom can be used to guide teaching and learning as it happens. The knowledge and skills required for the design and analysis of tests may not be readily transferable to this very different assessment approach.

The debate between the contrasting cultures of psychometric testing and sociocultural assessment is often characterised by a lack of mutual sympathy and can sometimes be decidedly hostile. From the sociocultural standpoint, van Lier (2004, 29–30) described standardised tests as “dehumanizing and oppressive pseudo-assessments” that are “linguistically flawed, morally indefensible and pedagogically harmful”. Fulcher (2012, 116–7), in contrast, considered the sociocultural approaches that van Lier advocates to be coercive. He considers that they require teachers to accept a relativist worldview that “reduces meaning to individual and group perception”, leaving no basis for judging the validity or value of assessment procedures.

19. Language assessment in practice

If the two paradigms cannot be reconciled, it might appear preferable that they should run on parallel tracks: courses in classroom assessment for teachers as facilitators of learning; courses in testing for those wishing to become specialists in the field working for testing agencies. Day-to-day decisions about how best to foster learning should involve assessment by teachers; high-stakes decisions that affect life chances should be informed by external testing. However, experience and research in the classroom suggest that in practice effective teaching implies coping with the competing demands of the full range of assessment purposes. Therefore, everything should be covered in training courses.

Rea-Dickins (2001, 432), based on classroom observation and interviews with teachers, argued that teachers would need to develop a variety of "identities of classroom assessment". These included a *bureaucratic identity* concerned with the more summative, externally driven demands for accountability; a *pedagogic identity* that monitors learners' progress and generates data that can be shared with other teachers; and a *learning identity* that focuses on the formative function of assessment as a means of learning and the role of the learner in this process. Each of these identities implies engagement with different forms of assessment literacy and the need to bridge the divide between them.

Teachers need to cope with externally mandated tests and accountability systems as well as assessing learner development. So, even those who favour a sociocultural approach towards training teachers may see the benefit of informing them about traditional testing as well. Scarino (2013, 312) suggested that "to address the range of assessment purposes, teachers of languages need to understand the assumptions of both paradigms and move between them." For this reason, Inbar-Lourie (2008) rejected Brindley's (2001) suggestion that teachers should initially study a module on *assessment in the language curriculum* and be exempted from *constructing and evaluating language tests* (see Appendix A).

Teachers who seek to foster learning through their classroom assessment practices will not usually escape the influences of large-scale external testing. They will need to develop appropriate competences to participate in and contend with the prevailing testing cultures that surround and constrain their practice. They need to find ways to reconcile their beliefs about teaching and learning with the requirement to report scores and grades, and the need to prepare learners for high-stakes tests. The contrasting paradigms approach to assessment literacy, therefore, involves awareness of and critical engagement with diverse traditions of assessment, even if they are seen to be mutually incompatible.

20. Ways forward

For all the antagonistic rhetoric, at a technical level, developing effective assessment procedures for the classroom would seem to have a good deal in common with developing effective tests. All forms of assessment, from informal questioning techniques to the multiple choice questions used in international examinations, benefit from attention to the social context and the purposes that assessments embody; theoretical insights into the nature of the abilities being assessed; shared critical review of assessment content and procedures; trialling the aforementioned;

and the careful analysis of results. The skills learned from making, sharing, and justifying judgements about performance on a test can also serve to sensitise teachers to learners' strengths and weaknesses in the classroom. Well-crafted test tasks reveal what learners know in ways that can also make them useful as tools for teaching and learning.

Whether highlighting commonalities or contrasts between the practices of classroom assessment and testing, commentators seem to agree that classroom assessment needs a stronger theoretical underpinning and a more extensive research base. In recent years, attempts have been made to build coherent theories, establish criteria for quality, and set an explicit agenda for research in this area (see for example Davison and Leung 2009; Hill and McNamara 2011). There would seem to be much to learn from the educational measurement tradition that informs language testing in these respects.

At the same time, large-scale language testing has been reforming itself to better match test content with current approaches to teaching and learning. Test developers have responded to demands for more authentic tasks that involve realistic performances in keeping with communicative or task-based approaches to teaching. Taking account of the increasing interest in test use and consequences, providers of major global language tests like the Educational Testing Service (US) and Cambridge English Language Assessment (UK) have been offering increasingly informative score reports that include some feedback on test-taker performance. They have built up their support for teachers preparing learners to take their tests. Resources such as online practice tests that provide instant results and test preparation guides are now widely available. Both organisations have implemented major projects to investigate the impact of their tests on language classrooms (Stoynoff 2012).

Such developments suggest the possibility of a complementary relationship between more formative and more summative assessment functions. Boud (2000, 160) suggested that "every act of assessment we devise or have a role in implementing has more than one purpose". He argued that assessments should be developed to perform what he called "double duty". This involves focusing both on the performance of the task at hand and the long-term educational goals: simultaneously indicating achievement of objectives and providing insights into the learning process. Assessment tasks that are designed or implemented with only one function in mind risk undermining the coherence of the educational process. Through washback, summative assessments may come to supplant important learning goals. The reason is that passing tests and gaining certificates seem to be more prized than developing a comprehensive